

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 04299399 A

(43) Date of publication of application: 22.10.1992

(51) Int. Cl G10L 3/02
G10L 7/08

(21) Application number: 03089601

(71) Applicant: RICOH CO LTD

(22) Date of filing: 28.03.1991

(72) Inventor: YASUDA HARUTAKE

(54) VOICE FEATURE QUANTITY EXTRACTION DEVICE

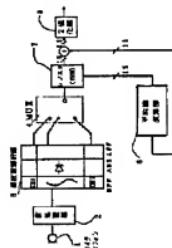
(57) Abstract:

PURPOSE: To provide the feature quantity extraction device which can normalize sound source characteristics through simple arithmetic when features are extracted by a BTSP system.

CONSTITUTION: The voice signal inputted from a microphone 1 is amplified and corrected by a preprocessing part 2. A frequency analysis part 3 obtains power spectra(TSP), channel by channel, and they are put into a time series by a multiplexer(MUX) and stored in a register 5 for 15 channels. The power spectra stored in the register 5 are supplied to a mean value arithmetic part 6 and mean values of three adjacent channels of the power spectra are found. An adder 7 subtracts the mean value of each channel, found by the

mean value arithmetic part 6, from the original power value to obtain the normalized sound characteristics of the speaker, thereby obtaining nearly the same effect with the least square approximate straight line.

COPYRIGHT: (C)1992,JPO&Japio



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平4-299399

(43)公開日 平成4年(1992)10月22日

(51)Int.Cl.⁵

G 1 0 L 3/02
7/08

識別記号 廈内整理番号

3 0 1 A 8842-5H
A 8842-5H

F I

技術表示箇所

(21)出願番号

特願平3-89601

(22)出願日

平成3年(1991)3月28日

(71)出願人 000006747

株式会社リコー

東京都大田区中馬込1丁目3番6号

(72)発明者 安田 晴剛

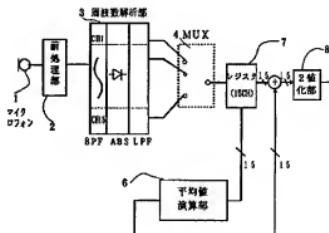
東京都大田区中馬込1丁目3番6号 株式会社リコー内

(54)【発明の名称】 音声特微量抽出装置

(57)【要約書】

【目的】 B T S P方式により特徴を抽出する場合に、より簡単な演算で音源特性を正規化することのできる音声特微量抽出装置を提供する。

【構成】 マイクロフォン1から入力された音声信号は、前処理部2で増幅、補正される。そして、周波数解析部3でパワースペクトル(T S P)が各チャネル毎に得られ、マルチブレクサ(MUX)4で時系列化され1 5チャネルのレジスタ5に格納される。レジスタ5に格納されたパワースペクトルは平均値演算部6に供給され、各チャネルのパワースペクトルに対して隣接する両側3チャネルの平均値が求められる。加算器5により、平均値演算部6で求めた各チャネルに対する平均値を元のパワースペクトル値から差し引けば、正規化された話者の音源特性が得られ、最小二乗近似直線とほぼ同等の効果を得ることができる。



〔特許請求の範囲〕

〔請求項1〕 入力された音声信号の周波数分析を行うnチャネルのバンドパスフィルタと、これらnチャネルの得られた各信号をパワーリミットと、この変換手段で得られるnチャネルのパワースペクトルから各チャネルの隣接するnチャネルの平均値を差し引いて正規化を行う正規化手段と、この正規化手段で得られたパワースペクトルを2値化する2値化手段とを具備することを特徴とする音声特微量抽出装置。

〔請求項2〕 入力された音声信号の周波数分析を行うnチャネルのバンドパスフィルタと、これらnチャネルの得られた各信号をパワーリミットと、この変換手段で得られるnチャネルのパワースペクトルの各ローカルピークに隣接するmチャネルの平均値のそれぞれのを結ぶ直線上の値を各パワースペクトルにより差し引いて正規化を行う正規化手段と、この正規化手段で得られたパワースペクトルを2値化する2値化手段とを具備することを特徴とする音声特微量抽出装置。

〔請求項3〕 入力された音声信号の周波数分析を行うnチャネルのバンドパスフィルタと、これらnチャネルの得られた各信号をパワーリミットと、この変換手段で得られるnチャネルのパワースペクトルの各ローカルピークの両方向に存在する極小値までの平均値のそれぞれを結ぶ直線上の値を各パワースペクトルから差し引いて正規化を行う正規化手段と、この正規化手段で得られたパワースペクトルを2値化する2値化手段とを具備することを特徴とする音声特微量抽出装置。

〔発明の詳細な説明〕

【0001】

【産業上の利用分野】 本発明は音声特微量抽出装置に係り、詳細には、フィルタバンク方式による音声認識において、フィルタバンクから得られるパワースペクトルの音源特性を補正して音声の特微量を抽出する音声特微量抽出装置に関する。

【0002】

【従来の技術】 人間と機械との対話を実現し各種O/A機器の操作を容易かつ迅速に行い、また工作機械の操作を安全に行うものとし、操作者が指示する内容の音声を認識し、その指示に従った処理や動作を行わせる音声認識技術が活発に行われ、商品化されるようになっている。音声の認識は、入力された音声の特微量を抽出し、予め登録してある標準パターンと比較することによって行われる。従って、音声認識を行う場合には、入力された音声からその特微量を抽出する方法がポイントとなる。このような音声の特微量を抽出する方法としてB T S P (Binary Time-Spectrum Pattern) 方式が知られている。B T S P方式は、T S P (Time-Spectrum Pattern) が声道の共振特性を示す伝達特性と音源特性を含み、音調固有の特徴を備えていることから、そのスペクトル上に現れるピーク（ローカルピーク）を2値化処理して特微量と

するものである。

【0003】 図4はB T S P方式による音声認識の概略のダイヤグラムを表したものである。マイクロフォンから入力された音声信号は、nチャネルのフィルタを備えたフィルタバンクで周波数分析され、パワースペクトルに変換される。このパワースペクトルから周波数上のピークを抽出し、それに基づいて「0」と「1」の2値化処理を行う。このように2値化したB T S Pを特微量を表す単語パターンとして取り扱い、未知の単語音声が入力されたとき、これを予め辞書に登録されている単語の標準パターンと照合し、認識結果を求める。

【0004】 ところで、母音などの音声を音源とする音声スペクトルは、高域で減衰していることが知られている。この減衰の個人性を補正する方法として、最小二乗近似直線を用いる方法が提案されている。B T S P方式でも2値化する前に、この操作により音源特性を正規化しておく。図5は、B T S P方式で最小二乗近似直線を用いて減衰を補正する処理過程を表したものである。ここでは便宜上、周波数軸でのみ説明する。まず、図5上段のT S Pの周波数軸上に、パワースペクトルの各点を基に次の式(1)で表される最小二乗近似直線(Least Square Fit Line)を引く。この式(1)で、yは周波数を表し、xはレベルを表す。

$$y = ax + b \quad \dots \dots (1)$$

その後、各ローカルピークを検出してピーク点近傍を「1」、他を「0」で表して、図5下段に示すような「0」と「1」からなる2値化パターンを生成する。このようにして、話者の音源特性は正規化され、得られた2値化パターンから音声の認識が行われる。

【0005】

【発明が解決しようとする課題】 しかしながら、この音源特性の正規化のための最小二乗近似直線の演算が非常に複雑となり、処理の高速化の点で問題があった。そこで、演算処理の簡便化のため、直線位相型C I C (Cascaded IntegratorComb) フィルタを用いた方が提案されている。この方式の概略を図6により説明する。音声のパワースペクトルを低域から高域、高域から低域とスキャンして時系列信号に変換し図6上段の時系列信号を得る。これを周期信号とみなして低周波成分前記フィルタで除去し、図6下段の音声スペクトルを得る。このフィルタは図4に示す構成のものが使用される。このC I C フィルタを使用した場合には、最小二乗近似直線による正規化よりも演算量が減少するが、以前として演算が複雑であるという問題がある。そこで本発明は、B T S P方式により特徴を抽出する場合に、より簡単な演算で音源特性を正規化することのできる音声特微量抽出装置を提供することにある。

【0006】

【課題を解決するための手段】 請求項1記載の発明では、入力された音声信号の周波数分析を行うnチャネル

3

のバンドパスフィルタと、これらnチャネルの得られた各信号をパワーに変換する変換手段と、この変換手段で得られるnチャネルのパワースペクトルから各チャネルの隣接するmチャネルの平均値を差し引いて正規化を行う正規化手段と、この正規化手段で得られたパワースペクトルを2値化する2値化手段を音声特微量抽出装置に具備させる。請求項2記載の発明では、変換手段で得られるnチャネルのパワースペクトルの各ローカルピークに隣接するmチャネルの平均値のそれぞれのを結ぶ直線上の値を各パワースペクトルにより差し引いて正規化を行う。請求項3記載の発明では、変換手段で得られるnチャネルのパワースペクトルの各ローカルピークの両方向に存在する極小値までの平均値のそれぞれを結ぶ直線上の値を各パワースペクトルから差し引いて正規化を行う。

【0007】

【実施例】以下、本発明の音声特微量抽出装置における一実施例を図1ないし図3を参照して詳細に説明する。図1は音声特微量抽出装置の構成を表したものである。図1において、1は認識する音声が入力されるマイクロフォン、2は音量の増幅や補正を行なう前処理部、3は音声信号を周波数解析しパワースペクトル(TSP)を得る周波数解析部である。周波数解析部3はc1h1からc1h1までの15チャネルのバンドパスフィルタ(BPF)、整流を行うABSおよびローパスフィルタ(LPF)から構成されている。4はパワースペクトルを時系列化するマルチブレクサ(MUX)、5は時系列化されたパワースペクトルを格納する15チャネルのレジスタである。6はパワースペクトルの平均値を求める平均値演算部、7はBTSPを得る加算器、8はBTSP2値化する2値化部である。

【0008】このように構成された音声特微量抽出装置における第1の実施例の動作について次に説明する。図1において、マイクロフォン1から入力された音声信号は前処理部2で增幅および補正された後に周波数解析部3に供給される。周波数解析部3では、供給された音声信号を15チャネルのBPFで周波数解析し、ABSで整流した後LPFを通してパワースペクトル(TSP)を得る。これら周波数解析部3で得られたパワースペクトルは、マルチブレクサ(MUX)で時系列化され15チャネルのレジスタ5に格納される。このようにして得られたパワースペクトルは、図2の実線で書かれたようなローカルピークを持ったスペクトル形状を一般に示している。

【0009】レジスタ5の各チャネルに格納されたパワースペクトルは平均値演算部6に供給され、各チャネルのパワースペクトルに対して隣接する両側mチャネルの平均値が求められる。この実施例では隣接する両側3チャネルの平均値を求めており、その値を図2の点線で示す。この方法は、いわゆる移動平均を求めたものと同様

で、LFPの効果があり、図のように時系列変換されたパワースペクトルの低周波成分のみが得られる。この出力は次の式(2)で求められる。この式(2)で、加算対象kの加算範囲は-mからmまでである。

$$y(n) = \Sigma x(n-k) \quad \dots \quad (2)$$

【0010】加算器5により、平均値演算部5で求めた各チャネルに対する平均値を元のパワースペクトル値から差し引けば、正規化された話者の音源特性が得られる。最小二乗近似直線とほぼ同等の効果を得ることができる。この出力を2値化部に送り、従来と同様にしてBTSPパターンを得ることになる。このようにして得られた特微量としてのBTSPパターンは、標準パターンとして図示しない辞書に登録される。また、同様にして得られた特微量は、同じく図示しない比較部で、予め登録された標準パターンと線型照合して類似度が算出され、最も類似する標準パターンの音声であると認識される。以上説明した第1の実施例では、平均化する場合のmの値を3としたが、mの値は任意の値でよいが、m=3~5の場合特に良好な結果が得られる。

【0011】次に第2の実施例について説明する。図3に、第1の実施例と同様にして得られたパワースペクトル信号を示す。この実施例では、15chのレジスタ5から供給されるパワースペクトルのローカルピーク点に着目し、ローカルピーク点のみ隣接するmチャネルの平均値を平均値演算部5で求める。そして、各ローカルピークの各平均点を結び、それを基に、元のパワースペクトル信号から加算器7で差し引いて音源特性を補正する。これらの2点間を結ぶ直線 $y=a: x+b$ 、 $y=a: x+b_1$ は幾何学的な演算により容易に計算することができます。

【0012】更に、第3の実施例について説明する。この実施例では、隣接するmチャネルの平均を求めるかわりに、各ローカルピークの両側の極小値となるチャネルまでを平均してその各ローカルピークの平均点を求めるが、各ローカルピーク単位の平均が求まり、同様にその平均点を結んで近似直線を得る。特殊な場合として、ローカルピークが1個しかなかった場合は、その平均点を基準としてそれ以下の部分を差し引けばよい。

【0013】

【発明の効果】本発明の音声特微量抽出装置によれば、音声の特微量抽出を簡便に且つ高速に行なうことができる。

【図面の簡単な説明】

【図1】本発明の音声特微量抽出装置における実施例の構成を示す構成図である。

【図2】音声特微量抽出装置の第1の実施例による正規化を説明するための説明図である。

【図3】音声特微量抽出装置の第2の実施例による正規化を説明するための説明図である。

【図4】BTSP方式による音声認識の流れを説明する説明図である。

50

5

6

【図 6】B T S P 方式で最小二乗近似直線を用いて減衰を補正する処理過程を説明するための説明図である。

【図 6】直線位相型 C I C フィルタを用いて減衰を補正する処理過程を説明するための説明図である。

【図 7】C I C フィルタの構成図である。

【符号の説明】

1 マイクロフォン

2 前処理部

3 周波数解析部

4 マルチブレクサ

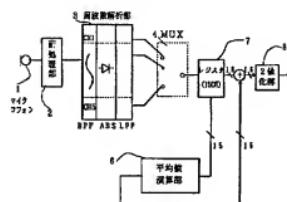
5 レジスタ

6 平均演算部

7 加算器

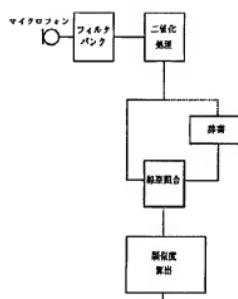
8 2 倍化部

[図 1]



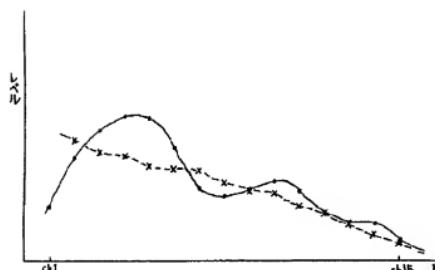
[図 2]

[図 4]

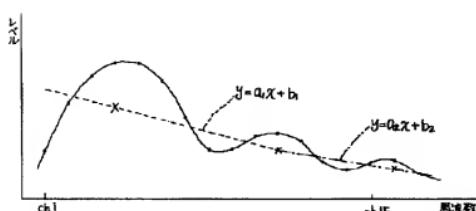
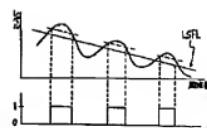


点線はトレーニングを示す

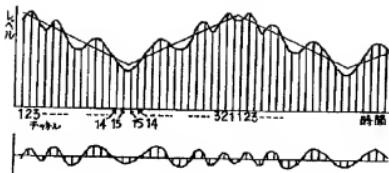
[図 5]



[図 3]



【図6】



【図7】

